知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于机器学习的我国城市“双创”水平的评价研究

◎ 史梦怡[1] 沈云骢[2] 李姣姣[1] 张思维[1] 徐梦宇[1]
[1] 复旦大学文献信息中心 上海 200433
[2] 复旦大学数学科学学院 上海 200433

摘要：[目的/意义]依托科技创新相关开放数据，针对全国城市“双创”水平建立评价体系与指数。[方法/过程]运用机器学习算法建立国内城市的“双创”排名，并通过各项指标和城市集聚特征分析城市“双创”能力优劣。[结果/结论]我国城市的“双创”水平与经济资源、人才储备、政策环境等有着密切的联系。北京市持续领跑中国“双创”发展，各区域中心城市紧随其后发展，各有优势，发展差距小。各城市的不足与短板也十分清晰，如何针对各自的不足做出合理的应对措施，扬长补短，是持续发展城市创新力的根本。

关键词：“双创”水平 创新创业能力 评价体系 “双创”指数

分类号：F272.2

引用格式：史梦怡，沈云骢，李姣姣，等. 基于机器学习的我国城市“双创”水平的评价研究[J/OL]. 知识管理论坛，2019, 4(2): 98-109[引用日期]. http://www.kmf.ac.cn/p/167/.

“大众创业、万众创新”被视作中国新常态下经济发展“双引擎”之一，它的提出旨在通过创新推动创业、创业带动创新的模式转换经济发展方式，促进经济转型发展。如今，随着信息技术的进一步发展，获取信息方式的多样化，使得创业成为每个普通人触手可及的发展可能，为解决就业问题提供可能；在创新驱动发展的战略背景下，各地区也为实现“双创”做出大量实践，以提高地区“双创”竞争力。然而地区之间的发展千差万别，存在的问题也不尽相同，因此，建立合理有效的“双创”评价指标体系显得尤为重要。

笔者通过对中国各城市“双创”环境的数据分析和解读，提出并建立“大众创业、万众创新”评价指标体系，以期指导中国各城市在“双创”活动的开展，针对活动合理性、科学性进行评估，发扬优势，弥补不足，促进创新活动的可持续性发展，为活动增效添质。

1 国内外创业、创新评价体系发展概况

国内外针对创业、创新评价体系业已展开

作者简介：史梦怡（ORCID：0000-0002-8465-8103），硕士研究生，E-mail：17210830021@fudan.edu.cn；沈云骢（ORCID: 0000-0002-1013-3443），博士研究生；李姣姣（ORCID：0000-0002-9230-6933），硕士研究生；张思维（ORCID：0000-0002-3645-1693），硕士研究生；徐梦宇（ORCID: 0000-0003-3611-6585），硕士研究生。

收稿日期：2018-08-13 发表日期：2019-04-19 本文责任编辑：刘远颖

ChinaXiv合作期刊
知识管理论坛，2019(2):98-109
DOI：10.13266/j.issn.2095-5472.2019.011

了一系列的研究，也拥有了相对成熟的体系。从目前来看，这种评价体系仅针对城市能力的某一方面建立，创业与创新的评价拆分成两个相互独立的方面。创业评价体系与创新评价体系独立发展，各自搭建起完整严密的评价体系。

1.1 创业评价体系概述

1.1.1 国外创业评价体系

从创业评价体系来说，国外相关的评价体系主要有 3 个：全球创业观察、全球创业与发展指数、考夫曼创业活动指数[1-4]。

全球创业观察（Global Entrepreneurship Monitor，GEM）是由英国伦敦学院与美国百森学院共同发起的国际创业研究项目，通过对全球各国家创业环境、创业项目、创业成果等的调查分析，发布《全球创业观察报告》，在全球范围内享誉盛名。全球创业观察所建立的评价体系，是立足于创业环境对创业活动的影响，通过对创业环境的构成要素进行拆分，从不同的环境角度进行深入分析，建立基于环境的创业评价体系；全球创业观察提出的是三层次环境要素理论，将创业环境拆分成 3 大要素：基础要求要素、效率提升要素和创新创业要素，其中包含金融支持、政府政策、政府项目支持、教育与培训、研究开发转移、商业和专业基础设施、进入壁垒、有形基础设施、文化与社会规范等 9 个方面，其创业评价体系具有一定的广度与深度，因此《全球创业观察报告》极具影响力[1,5]。

全球创业与发展指数（Global Entrepreneurship and Development Index，GEDI），其产生主要是为了更加全面立体地了解在地区发展中，促进企业成立和成长壮大的典型环境特征，进而对该地区高效率创业支持程度进行评估。GEDI 指标体系包含了创业态度、创业活动、创业愿望 3 个子维度，该指标体系囊括了地区发展体系中促进这 3 个子维度的 34 个最主要的制度因素和个人因素，能够预测该地区的创业潜能[2,5]。

考夫曼创业活动指数（Kauffman Index of Entrepreneurial Activity，KIEA）是美国早期具有重要参考价值的创业活动指数，其利用人口指数，计算成人非企业主人口中新创办企业的人口比例，来确定创业活动指数，对新创企业的首月商业活动密切关注[3]。

1.1.2 国内创业评价体系

国内较为综合的创业评估体系[6-8]有中关村指数、城市创业环境评价指标体系、中国城市创业指数等。

中关村指数是由北京市统计局编制的，是用以观察北京市高新技术产业发展状况的综合指数，包括经济增长指数、经济效益指数、技术创新指数、人力资本指数和企业发展指数 5 个大分类指数，是一个更具宏观角度的综合性评价体系。

城市创业环境评价指标体系则侧重创业环境的优劣，从环境硬指标到环境软指标，即能源、邮电通讯、市政、制度、管理、人文 6 个角度评价城市创业环境。

1.2 创新评价体系概述

1.2.1 国外创新评价体系

从创新评价体系来说，目前国际上主要有 4 个创新评价体系，全球创新指数（Global Innovation Index，GII）、欧盟创新指数（Summary Innovation Index，SII）、硅谷指数（Joint Venture Silicon Valley，JVSV）和全球知识竞争力指数（World Knowledge Competition Index，WKCI）[5]。

全球创新指数通过评估制度和政策、创新驱动、知识创造、企业创新、技术应用与知识产权等，为企业领袖与政府决策者提供提升综合竞争力可能面临的缺失与努力方向，对一个经济体广泛的经济创新能力进行衡量。

欧盟创新指数经过多年的指标调整，从人力资源、金融支持、企业投资、联系与创业、生产率、创新、经济效果几个方面，定量表现地区的创新现状，分析地区之间的现状差异。其创新指数体系主要包括欧洲创新记分牌、区域创新记分牌和创新晴雨表调查。

硅谷指数指标体系则侧重于从宏观角度对创新基本条件进行衡量，包括人口、经济、社会、空间和管理指标。

全球知识竞争力指数主要是针对知识经济做出的指标体系，主要指标为人力资本、知识资本、金融资本、经济产出和知识可持续性，但未对创新产出设立指标。

1.2.2 国内创新评价体系

国内创新评价体系主要集中在有关区域科技发展层面，在此过程中也借鉴了众多国外相关经验，建立起将“科技创新能力要素”作为主要评价标准的一整套评价机制。科技创新能力要素主要通过对区域科技的投入和产出进行分析，根据区域科技创新的投入指标和区域科技创新的产出成果指标来建立评价体系。关于科技创新投入指标及产出成果，通常根据基础环境、科技投入、科技产出、科技成果转化、经济发展、创新效益等指标来建立科技创新评价体系。

中国科技发展战略研究小组[9]提出从知识创造、知识流动、企业技术创新能力、创新环境和创新的经济绩效 5 个方面评价区域创新能力，较早形成了比较权威的区域创新能力评价指标体系。

1.3 研究述评

从国内外创新创业体系来看，国外在城市能力评价体系上的研究更为深入，指标的选取和建立也更加成熟，国内创新创业评价体系则存在一定的局限性，内容更多偏向于特定领域。

综合分析各创新创业评价体系后发现，其指标的建立主要集中在政策环境、经济实力、投入产出比、人才发展等众多宏观层面，涵盖的内容较为充足，所以笔者综合国内外创新创业评价体系，同时结合国务院在“大众创业，万众创新”的实施意见中对创业环境、创业活力、创新创业政策等方面提出的建设思路，建立“双创”评估体系，针对政策环境，列举政府投资、基础设施建设等细分指标；针对经济实力，列举 GDP 增长、公司利润增长等细分指标；针对投入产出比，列举研发支出与占比、专利产出等细分指标；针对人才发展，列举高等院校数量、科研机构数量等细分指标；基于此，对“双创”能力评价方法进行探讨，意在为我国各地区“双创”评估提供具有价值性的参考。

2 数据来源与处理

2.1 数据来源

本文数据主要来源于国家统计局、中国经济与社会发展统计数据库、Wind 数据库、中华人民共和国教育局、中国人民银行以及发布于北京大学开放研究数据平台的国信宏数的开放数据库[10-13]。

国家统计局数据库与中国经济和社会发展统计数据库可以说是目前中国最权威的官方统计数据库。笔者从这两个数据库中收集了人口、GDP 等宏观测度以及各项反映城市基础设施建设水平的数据，并以来自中华人民共和国教育局公布的高校及“双一流”名单和中国人民银行公开的金融数据作为补充。

Wind 数据库是以金融证券数据为核心的一流大型金融工程和财经数据仓库。笔者从 Wind 数据库中选取了上市公司财报数据中研发支出、就业者学历水平、技术人员占比、行业分布以及利润和营业额水平等指标，作为反映我国创新创业政府、经济、人才等方面资源实力的数据支撑。

发布于北京大学开放研究数据平台的国信宏数的开放数据库是专注于宏观经济大数据分析的数据库，涵盖了发改、工商、税务、金融等领域的大量数据。笔者从中选用了首届全国高校数据驱动创新研究大赛官方数据空间的多组数据，包括“地图兴趣点（Point of Interest，POI）数据”“互联网招聘网站数据”“发明专利数据”等，开展“双创”水平评价体系的建设。

2.2 数据处理

数据处理包括数据清洗、数据预处理等步骤。数据清洗是指删除、更正数据集中错误、不

知识管理论坛，2019(2):98-109
DOI：10.13266/j.issn.2095-5472.2019.011

完整、格式有问题或多余的数据。数据清洗主要是对缺失数据和异常数据进行处理。对于缺失数据，采取加权平均法和插值拟合法对数据空值处进行填充。而后，在对清洗完成的数据进行统计分析时，发现常有极端值、不连续值等异常数据存在，于是按照实际情况进行修正或加权平均，消除异常数据对结果的影响。最后，将数据正则化，作为"双创"水平评估的各项指标值。

利用Excel与MATLAB等工具，对各来源、各类型的数据进行了处理与分析。"地图兴趣点POI数据"是采自2017年夏高德地图兴趣点的数据。"互联网招聘网站数据"是反映城市人才水平的重要参考依据，而其中记录条数、岗位数、各学历人员占比都是重要的小指标；而又由于互联网招聘网站数据的非权威性以及城市发展偏差，常常可能出现极端值情况，需要特别注意。相比之下，"发明专利数据"的数据可靠性较高，统计处理也相对简单，只需考虑发明、实用新型和外观设计3类专利的含金量水平的差异。

除此之外，笔者通过加权平均法，根据过去若干个按照时间顺序排列起来的同一变量的观测值并以时间顺序数为权数，计算出观测值的加权算术平均数，以这一数值预测未来该变量变化趋势，例如，以各省的国民生产总值（GDP）为基础，首先计算出各地级市国民生产总值（GDP）占该省的比例，然后再以该比例为权重计算各省市其他地级市的各项指标值。其中按照加权平均法计算的指标有：城市道路面积、电力消耗量、供水总量。

3 "双创"指标体系构建

3.1 指标体系建立

本文的研究对象为2018年国家统计局数据库中统计到GDP及人口数据的城市及地区，包括直辖市和自治州等。综合已有文献[14-16]，系统考虑政策、经济、人才、基础设施等因素，以发表专利数表示"双创"产出，笔者最终确定主体、资源、环境、绩效4个层面和"双创"能力、"双创"信心、政府资源、经济资源、人才资源、基础设施、经济水平、环境氛围、专利产出等12个因素作为"双创"水平评价指标体系，如表1所示：

表1 "双创"水平评价指标体系

层面	因素	指标
主体	"双创"能力	高等院校数
		万人拥有高等院校数
		重点以上高校数量
		国家级科研机构数量
	"双创"信心	上市公司研发支出占比
资源	政府资源	上市公司政府投资额
	经济资源	社会融资增量
		研发支出额
	人才资源	上市公司技术人员占比
		互联网招聘人才水平
环境	基础设施	供水总量
		电力消耗量
		城市道路面积
	经济水平	GDP近三年均增长率
		上市公司近三年均利润增长
	环境氛围	地图POI文娱占比量
		地图POI科研占比量
		地图POI教育占比量
绩效	专利产出	专利申请授权数

将所收集数据按指标进行整理，最终得到了城市各分项"双创"指标值。但由于各项数据量纲不统一，且数据质量参差不齐，偶有数据两极分化严重，不便于后续操作。故利用MATLAB，以原数据各列的最大值、中位数和四分位数为参考点，按照公式：，拟合幂次公式参数 a,b,c，再对数据进行正则化，提高数据的平滑性和整齐度。其中，d_i 为原始数据，d_i' 为正则化后的数据，a,b,c 为幂次公式模型参数。最终，得到全国城市的"双创"各项指标得分，如表2所示：



知识管理论坛
2019 年 第 2 期(总第 20 期)

表 2 全国城市的“双创”各项指标得分

城市			北京	上海	深圳	南京	成都
主体	“双创”能力	高等院校数	100.00	95.74	73.39	93.59	94.21
		万人拥有高等院校数	34.30	21.16	16.48	70.02	39.66
		重点以上高校数量	100.00	80.00	0.00	76.76	66.06
		国家级科研机构数量	100.00	100.00	100.00	100.00	100.00
	“双创”信心	上市公司研发支出占比	17.11	59.89	61.62	65.46	12.50
资源	政府资源	上市公司政府投资额	100.00	95.34	90.76	85.28	82.01
	经济资源	社会融资增量	95.62	100.00	91.93	84.94	85.96
		研发支出额	100.00	94.51	93.67	84.28	82.87
	人才资源	上市公司技术人员占比	82.12	59.89	61.62	65.46	66.84
		互联网招聘人才水平	100.00	50.68	17.95	22.33	27.05
环境	基础设施	供水总量	87.23	65.64	69.42	54.96	75.19
		电力消耗量	81.45	59.93	65.66	48.85	72.01
		城市道路面积	78.79	57.98	34.17	74.95	50.66
	经济水平	GDP 近三年均增长率	74.79	74.69	74.88	74.90	74.87
		上市公司近三年均利润增长	71.97	72.89	73.24	73.58	72.70
	环境氛围	地图 POI 文娱占比量	100.00	89.53	81.17	86.47	85.80
		地图 POI 科研占比量	100.00	78.15	69.17	77.30	78.14
		地图 POI 教育占比量	100.00	74.80	66.66	73.18	79.07
绩效	专利产出	专利申请授权数	100.00	95.31	96.11	89.32	90.27

3.2 指标权重考量

上一节确定的“双创”水平评价体系综合考虑了主体、资源、环境、绩效 4 个层面、12 项因素，共产生 19 项指标得分。而如何衡量每项指标的重要性，并综合得到城市“双创”水平总分，就涉及到了权重和算法的考量。

具体来说，每个城市 i 都有一个指标 x_i，它是一个 19 维的 w 向量，我们希望找到一个权重向量，使得它们的内积 (x_i,w) 能反映最后的综合得分。

现阶段，已有许多组织和机构按照自己的评价体系，对全国重点城市的“双创”水平进行了评分和排名，包括《第一财经》《国家治理周刊》《中国省区、直辖市综合竞争力评价指标体系》和复旦经济学等，具有重要的参考价值。

笔者首先以上述机构所公布的全国重点城市的排名和指标的得分作为机器学习的训练集，分别是北京（99.422）、深圳（93.296）、上海（89.416）、广州（78.565）、苏州（78.518）、武汉（75.353）、杭州（74.991）、南京（74.893）、天津（74.319）、成都（70.966）、西安（67.097）、长沙（63.500）、无锡（63.500）、重庆（61.869）、青岛（61.673）、合肥（60.241）、宁波（58.683）、济南（57.953）、厦门（55.567）、大连（54.477）、哈尔滨（49.722）、沈阳（48.395）和长春（46.792），以前一节构建的指标体系为因子，在 MATLAB 中的用支持向量机和最小二乘两种方法进行回归。

其中支持向量机采用的是线性核，这样训练出来的系数就是权重，据此再对所有的城市计算得分和排名。但结果与理想情况偏差较大，因子间权重相差极大，多数情况只有一两

知识管理论坛，2019(2):98-109
DOI: 10.13266/j.issn.2095-5472.2019.011

项指标在评分体系中有作用。

另外，这两种方法得到的系数并不一定是正的，由于训练样本都是得分较高城市，如果出现负的权重，会使得样本外某些该项得分低的城市排名出现异常上升的情况。因此，对于所采用的数据，负权重是不合理的。

于是需进一步改进机器学习的优化目标。首先，限定系数必须是非负的，并根据评价需求与实际情况，设置一定的上下界。其次，对目标中每个数据的权重进行调整，分值越高的权重越大。这样操作之后效果有明显提升，最终笔者选取如下的优化模型：

$$\min_{w\geq 0}\sum_{i\in T} y_i^2(x_i\cdot w-y_i)^2$$

其中，T 表示训练样本集，y_i 表示样本得分。这是一个二次优化问题，算出权重系数 w，如表 3 所示：

表 3 "双创"水平评价指标体系各项权重

层面	因素	指标	权重系数
主体	"双创"能力	高等院校数	8.3687E-23
		万人拥有高等院校数	3.9369E-22
		重点以上高校数量	7.2628E-20
		国家级科研机构数量	6.5246E-02
	"双创"信心	上市公司研发支出占比	6.3006E-02
资源	政府资源	上市公司政府投资额	5.5049E-22
	经济资源	社会融资增量	3.6054E-01
		研发支出额	2.3820E-01
	人才资源	上市公司技术人员占比	1.1276E-01
		互联网招聘人才水平	1.1347E-01
环境	基础设施	供水总量	2.4169E-22
		电力消耗量	2.0990E-22
		城市道路面积	1.1709E-18
	经济水平	GDP 近三年均增长率	4.5566E-23
		上市公司近三年均利润增长	1.1674E-23
	环境氛围	地图 POI 文娱占比量	4.7388E-23
		地图 POI 科研占比量	1.0000E-01
		地图 POI 教育占比量	1.5560E-22
绩效	专利产出	专利申请授权数	2.1675E-22

从权重系数来看，社会融资增量、研发支出额和上市公司研发支出占比作为创新创业经济资源水平的体现，对城市"双创"水平的提升具有重要意义，而上市公司技术人员占比和互联网招聘人才水平也为城市创新创业提供了人才支撑，国家科研机构数量和地图 POI 科研机构的占比量反映了城市的"双创"能力和"双创"氛围，在评价体系中均占有重要的地位。

相较而言，从权重角度来看，环境因素中基础设施、经济水平等对城市"双创"能力的影响较小；在基础设施中城市道路面积作为交通指标的代表，相比供水总量和电力消耗量更能对"双创"能力起到正面影响。

4 分析及结果讨论

4.1 "双创"指数评分结果分析

从经济、政治、人才教育、社会环境等多个角度出发，依据上一章所建立的数据模型，综合评估中国各大城市的创新创业能力，最终得出中国城市创新创业能力排名及得分，具体如表 4 所示：

表 4 中国城市创新创业能力排名及得分

排名	城市	得分
1	北京市	96.91
2	上海市	86.88
3	深圳市	77.80
4	南京市	77.43
5	成都市	77.26
6	天津市	76.57
7	郑州市	75.30
8	武汉市	73.15
9	杭州市	72.59
10	无锡市	71.16
11	苏州市	69.86
12	广州市	68.75

从"双创"城市区域分布来看，东西部发展差异明显，沿海地区整体发展也优于内陆地区；长三角地区、京津冀地区、珠三角地区形



知识管理论坛
2019 年 第 2 期（总第 20 期）

成了明显的区域集聚趋势，“双创”的发展区域化显著。

由此可见，城市之间的辐射作用在“双创”能力发展中同样十分重要，3 个地区的三足鼎立格局稳定；长三角地区城市集聚数量最多，具有十分明显的优势；京津冀地区依托北京市、天津市形成稳固的发展格局；珠三角地区拥有深圳市、广州市两大“双创”领头羊，带动珠三角地区其他城市“双创”发展。围绕成都市、武汉市、郑州市，中西部区域均形成了小块城市集聚效应，城市的区域中心地位稳固，辐射作用明显。

根据综合得分可以看到，北京市、上海市在“双创”能力中具有明显的优势，处在“双创”城市的第一梯队，而紧随其后的深圳、南京、成都在综合得分上的差距很小，是“双创”城市的第二梯队，而包含了剩下城市的第三梯队与第二梯队之间的得分差距不大，在“双创”实力上同样具有竞争力。

从“双创”能力排名前 5 的城市各指标分布雷达图来看（见图 1），在政府投资方面，Top5 城市排名与总体排名吻合，可以说政府投资是“双创”能力的重要影响因素之一，科研创新产出在“双创”能力中也占有重要地位。

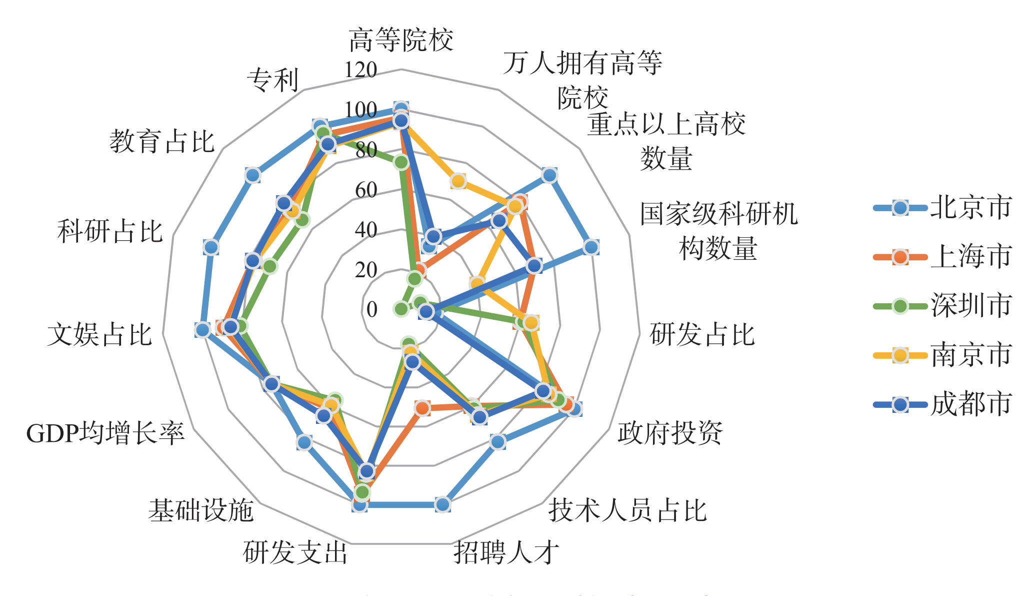


图 1 全国 TOP5 城市“双创”水平雷达图

北京市各方面优势均十分突出，仅在“万人拥有高等院校数量”“研发占比”上存在短板，资源分配是北京市面临的主要问题；上海、深圳、南京、成都在各项指标上的表现差距不大，各有千秋，上海市发展从总体上来说与北京相似，各项指标表现均衡，但同样面临教育资源合理分配的困境；深圳市在高等院校、科研机构数量上的劣势明显，但在研发支出、专利产出方面的表现突出，创新投入和成果转化工作相对完善；南京市的教育资源丰富，万人拥有高等院校数量排名第一，但国家级科研机构数量不多，科研实力略逊色，人才引进制度不够完善；成都市拥有众多国家级科研机构，从科研产出、占比方面来看具有一定优势，基础设施建设优于上海、深圳、南京，但其优势不够突出，整体表现相对均衡。

（1）TOP1 北京市。从整体排名和得分上可以看到，北京市作为中国的首都，其创新创业能力在中国各大城市中的表现尤为突出，这与北京市政治经济文化资源的强大优势是分不开的。高级人才资源储备丰富、“双创”人力、财力投入力度大是北京“双创”能力领先的主要

知识管理论坛，2019（2）:98-109
DOI：10.13266/j.issn.2095-5472.2019.011

因素。北京市的政府投资在中国各大城市中排名第一，达到463亿人民币，强有力的政府资金支持是北京市创新创业活动开展的基础；同时，北京市在人才储备方面的资源十分丰富，各大高校、科研所的聚集，使得北京市在创新创业方面拥有源源不断的人才供给。在资源丰富的基础上，优秀资源整合平台迅速崛起，国家级众创空间达125家，居中国第一，以创新工场、36氪、联想之星等为代表的创新型孵化器是其中最具活力的组成部分。

然而北京市面临的创新创业困境也十分明显，人口与资源环境之间的矛盾突出，资源分配成为北京市如今发展的一大掣肘，各项创新创业活动的成本上升，环境污染严重，政府资源分配紧张，都成为创新创业活动开展的阻力。

（2）TOP2 上海市。上海市的创新创业能力在中国各大城市中排名第二，虽与北京之间存在一定的差距，但也独具优势。作为中国最重要的金融商贸中心，上海市依托长三角城市群经济腹地，逐渐建立起自己的经济优势，在政策上，除了拥有排名第二的政府投资外，上海市还出台了一系列相关政策以支持上海市“双创”发展，为引进人才，制定发布海外人才引进、户籍政策以及国际人才试验区和职称制度改革等政策，创新创业人才环境得到持续优化，建立了多处“双创”示范基地。

上海的众创空间也蓬勃发展。截至2017年上半年，上海市共有众创空间500余家，其中创业苗圃100家，孵化器159家，加速器14家，创客空间等新型孵化器250余家以及科技型中小企业12 000多家。2016年举办各类创业辅导活动9 800多场次，覆盖38万余名创业者，为创业者和创业企业打造了低成本、便利化、全要素、开放式的综合创业网络和载体[14]。

而上海市在“双创”发展中的问题与北京市相似，高昂的资源成本、分配不均，本土排外情绪等都成为“双创”发展的挑战。

（3）TOP3 深圳市。深圳市作为中国改革开放后建立的第一个经济特区，近些年来呈现出了迅速发展的势头，世界瞩目；可以说，深圳市是一个年轻的城市，它在“双创”能力中的表现也同样表现出了年轻的活力。从基础设施建设和人才储备来说，深圳市并不具有优势，反而略显捉襟见肘，教育资源、人才资源是深圳市发展的主要短板；但深圳市在专利产出方面，却拥有能与北京市相媲美的实力，专利产出排名第二，表现亮眼，这主要归功于深圳市高科技企业的迅速发展；除了汇集以科技为主导的高科技公司外，众多国内领先的创投机构也在深圳市扎根成长。同时，深圳围绕“双创”需求构建政策链，制定促进创客发展的若干措施，安排多项“双创”资金，构建了覆盖不同创新主体和创新创业全过程的政策体系。

从长远来看，深圳市稀缺的教育基础资源是需要慎重考虑的问题，人才储备量不足，容易造成城市“双创”发展的后劲不足，同时，建立更为完善的人才引进制度，也是深圳市“双创”环境可持续发展的重要一步。

（4）TOP4 南京市。南京市是一座具有深厚文化底蕴和经济实力的城市，依托南京的人才储备优势以及江苏省“双创”政策的出台，建立了自己的“双创”特色发展道路，同时与无锡、苏州形成“双创”联动体，共同推进“双创”环境的建设和完善。

南京市在教育资源分配上拥有相当的优势，人才储备丰富，但从招聘人才数量来看，南京市人才流失情况严重，在人才保留措施上存在疏漏，因此，人才保留是南京市“双创”能力实现进一步突破需要重点关注的部分。

（5）TOP5 成都市。成都市位于中国中西部，是中西部城市发展的一颗明珠。“新一线城市”成都力压天津、杭州、广州等，在“双创”能力排名中一举拿下第五位的名次，这也受惠于成都当地政府的支持、人才和资金的储备及投入，基础设施建设完善。

而影响成都市“双创”发展的问题也十分清晰，中西部的地理位置使得成都在人才吸引方面仍要做出更多的努力，同时，虽然在研发

方面，成都市拥有相当的科研机构资源，但成果转化的效果并不明显，高新企业的稀缺也是成都市“双创”持续发展的一大难题。

除了以上 5 座城市外，天津、郑州、武汉、杭州、无锡、苏州、广州都拥有雄厚的“双创”实力，天津市拥有深厚的人才教育储备以及丰厚的政府投资支持，综合来看“双创”实力并不逊色；郑州市拥有良好的科研资源，创新科研实力有很好的保证，但从综合实力来说，郑州市的人才储备、政府投入方面都存在不可避免的短板；武汉市是除了成都市外的中西部“双创”发展重点城市，其人才储备方面的实力在各城市中排名前列，但在科研发展方面略显疲态；杭州市依托电子商务快速发展，为杭州创新创业活动注入了强大动力，使杭州迅速成为“创业新城”，政府支持力度大，发展活力足；无锡和苏州是除南京市外，江苏省“双创”发展的另外两只领头羊，其中苏州市在人才储备方面较无锡略胜一筹，“双创”青年数量全省排名第一，无锡夺得第三，在专利创新产出方面，苏州实力强劲；无锡市则在科研储备方面较苏州有明显优势，且无锡市更为注重人才引进，人才招聘数量全国排名第二；无锡市在各项指标中的表现更为稳定，发展均衡，而苏州在人才储备、科技创新产出方面的表现突出，其他方面的表现则相对逊色；广州市作为中国特大城市之一，其教育资源、政府投入的表现处在全国前列，但从人才引进、科研支出以及近三年的 GDP 利润增长的角度来说，广州市表现欠佳，人才储备充足，但人才流失成为广州市“双创”能力持续发展的掣肘。

4.2 全国城市“双创”空间集聚特征分析

除了城市自身的经济、人力、环境等水平有差异之外，不同城市的“双创”水平的空间关联特征也大有不同。何舜辉等在《中国地级以上城市创新能力的时空格局演变及影响因素分析》一文中研究发现，相邻城市的创新水平具有较高的一致性，在空间分布上呈现集聚现象，为了进一步反映局域空间关联特征，将不同局域的关联模式归纳为 4 类 [15]：① H-H 集聚区（高效型）城市本身的创新能力水平较高，且周围相邻城市的创新水平也较高，从而成为中国的“高效型”创新区域；② H-L 集聚区（极化型）城市相对于周围城市具有较高的创新能力水平，但无法带动周边地区创新能力的提高，从而形成了中间高、周围低的“极化型”的分化格局；③ L-H 集聚区（空心型）城市自身创新水平较差，且受周围城市的影响也小，创新能力明显低于周围城市，形成周围高，中间低的“空心型”的分化格局；④ L-L 集聚区（低效型）城市的创新能力属于国内较低水平，且与周围城市的差异较小，是创新产出的低值集中区。

利用 MATLAB，根据地理坐标信息，对城市及其周围城市的“双创”指数进行量化分析，可以清晰地观察这 4 类城市在全国的地区分布情况，具体见表 5- 表 8。

表 5 H–H 集聚区（高效型）城市分布

城市	“双创”指数	周边指数
北京市	96.909 32	55.935 53
天津市	76.570 21	59.883 71
苏州市	69.863 76	60.637 02
珠海市	66.498 74	63.399 73
沧州市	50.101 59	65.214 77
南通市	60.567 22	68.506 38
绍兴市	61.524 33	64.708 77
宁波市	63.270 25	61.843 16
上海市	86.876 02	58.883 29
南京市	77.426 06	56.326 1
无锡市	71.159 68	57.359 81
东莞市	58.188 02	66.695 27
保定市	60.693 85	60.148 37
中山市	59.175 89	61.557 11

H-H 集聚区（高效型）主要分布在长三角、环渤海和珠三角地区。不过在同一尺度下，京津冀地区北京、天津本身的“双创”水平较高，然而周边仍有廊坊市、承德市、张家口等“双创”能

知识管理论坛，2019(2):98-109
DOI：10.13266/j.issn.2095-5472.2019.011

力一般的城市，拉低了区域创新水平；相比之下，长三角地区上海、南京、杭州本身的“双创”水平较高，而周边绍兴、无锡、南通等城市的“双创”水平也均较高，充分发挥了城市集聚的优势，互通有无，高效合作。

表 6 H-L 集聚区（极化型）城市分布

城市	“双创”指数	周边指数
成都市	77.257 14	40.359 17
武汉市	73.153 71	30.519 84
许昌市	68.160 27	46.768 4
石家庄市	67.914 23	40.976 87
重庆市	67.730 68	30.131 26
西安市	67.514 35	17.406 07
长春市	66.732 44	32.218 25
沈阳市	66.353 39	47.795 92
十堰市	66.172 72	35.939 55
合肥市	65.997 23	42.994 51

H-L 集聚区（极化型）主要分布在内陆地区和东北地区的省会城市。例如，成都、武汉、重庆等中西部城市以及长春、沈阳、哈尔滨等东北省会城市，吸引了来自全省的人才、资金力量，表现出了不凡的“双创”能力，但对于周边城市的带动力较弱，还需要加强区域交流与合作，谋求共赢。

表 7 L-H 集聚区（空心型）城市分布

城市	“双创”指数	周边指数
廊坊市	38.596 03	67.598 19
舟山市	45.532 04	65.390 8
承德市	41.062 91	63.659 77
张家口市	40.523 02	62.229 23
泰州市	47.552 16	61.087 34
淮安市	46.797 54	60.790 99
滁州市	49.747 33	60.549 22
清远市	24.394 53	60.048 08
萍乡市	43.813 33	59.585 83

L-H 集聚区（空心型）呈零星点状分布于我国东部和中部地区，例如廊坊、衡水、舟山、黄山等城市。这些城市周边的创新水平整体较高，但其自身或许出于地理或经济条件的限制，无法与周围城市形成良好的合作。

表 8 L-L 集聚区（低效型）城市分布

城市	“双创”指数	周边指数
蚌埠市	48.343 41	49.544 47
衡阳市	48.354 01	49.203 29
漯河市	47.034 37	49.724 82
安庆市	47.396 93	49.274 36
开封市	48.220 23	46.631 98
淮北市	47.425 87	46.708 36
德阳市	45.882 58	47.972 55
邢台市	45.209 85	47.692 63
乐山市	47.188 24	45.294 73
丹东市	43.911 98	48.386 72
周口市	47.293 36	44.071 52
淮南市	43.418 02	47.918 4
晋中市	48.522 33	41.985 09

L-L 集聚区（低效型）城市广泛分布于中国内陆各地，例如四川的乐山市、山西的大同市和河南的开封市。直观来看，这些城市的“双创”水平整体低迷，还应从自身的经济、人力、环境等因素着手改善，提高“双创”能力。

5 结语

笔者依据国家统计局、中国经济与社会发展统计数据库、Wind 数据库、中华人民共和国教育局、中国人民银行以及发布于北京大学开放研究数据平台的国信宏数的开放数据库构建了“双创”水平评价指标，其中通过对互联网招聘大数据、城市地图 POI 大数据的应用，评价指标的真实性、时效性、丰富性相比传统评价指标有很大提高。在数据库内容持续更新的情况下，可通过笔者构建的指标对城市“双创”水平进行即时跟踪研究，加快决策反馈速度。

综合来看，“双创”实力与政府投资、人才储备、创新产出、政策环境等有着密切的联系，政府投资更倾向于中国的特大城市，北上

广在投资方面资源更为丰富，具有先天的基础优势；但各区域中心城市在完善政策环境方面不遗余力，在人才储备方面，加强人才引进，从智力资源、科研创新寻找突破口。

“双创”能力的强弱是一个综合评价的结果，而“双创”能力的可持续发展则是一个动态的过程。北京市持续领跑中国“双创”发展，各区域中心城市紧随其后，各有优势，发展差距小。与此同时，各城市的不足与短板也十分清晰，如何针对各自的不足提出合理的应对措施，扬长补短，是持续发展城市创新力的根本。我们期待各城市在保持优势的基础上，正确认识不足，及时采取相应的措施，全面提高城市创新创业竞争力，形成更繁荣的百花争艳景象。

参考文献：

[1] HAY M, COX L W, REYNOLDS P D, et al. Global entrepreneurship monitor (Gem)-2002 executive report[J]. Social science electronic publishing, 2002, 3(1):66–70..

[2] ACS Z J, SZERB L, AUTIO E. The global entrepreneurship and development index[J]. Regional studies, 2015, 49(12): 1977-1994..

[3] FAIRLIE R W. Kauffman index of entrepreneurial activity by veteran status 1996-2011[J]. Ssrn electronic journal, 2012(21): 28.

[4] CHEN X, LIU Z, MA C. Chinese innovation-driving factors: regional structure, innovation effect, and economic development—empirical research based on panel data[J]. The annals of regional science, 2017, 59(1): 43-68.

[5] 胡平．“大众创业，万众创新”评价体系与评价方法探讨 [J]. 科学与管理，2017, 37(5): 14-21.

[6] 周立，吴玉鸣．中国区域创新能力：因素分析与聚类研究——兼论区域创新能力综合评价的因素分析替代方法 [J]. 中国软科学，2006(8): 96-103.

[7] 何健文．广东区域创新能力分析报告——基于《中国区域创新能力报告 2015》[J]. 广东科技，2017, 26(2): 47-52.

[8] 白嘉．中国区域技术创新能力的评价与比较 [J]. 科学管理研究，2012, 30(1): 15-18.

[9] 中国科技发展战略研究小组．2003 中国区域创新能力报告 [M]. 北京：经济管理出版社，2004.

[10] 中华人民共和国国家统计局 [EB/OL]. [2018-01-25]. http://www.stats.gov.cn/.

[11] 中国经济与社会发展统计数据库 [EB/OL]. [2018-01-27]. http://tongji.cnki.net/kns55/index.aspx.

[12] 中华人民共和国教育局 [EB/OL]. [2018-01-27]. http://www.moe.gov.cn/.

[13] 国信宏数 [EB/OL]. [2018-01-27]. http://www.ndrcbd.org/.

[14] 臧鸣．双创促升级 壮大新动能丨上海已有 7 家全国双创示范基地 [EB/OL]. [2018-01-20]. http://www.thepaper.cn/newsDetail_forward_1794996.

[15] 何舜辉，杜德斌，焦美琪，等．中国地级以上城市创新能力的时空格局演变及影响因素分析 [J]. 地理科学，2017, 37(7): 1014-1022.

作者贡献说明：

史梦怡：数据处理，结果分析，论文结构搭建；
沈云骢：数据深度处理，模型设计；
李姣姣：数据收集及预处理，论文写作，论文修改完善；
张思维：文献综述的撰写以及论文写作；
徐梦宇：数据处理，模型设计。

ChinaXiv合作期刊
知识管理论坛，2019（2）:98-109
DOI：10.13266/j.issn.2095-5472.2019.011

Evaluation System on Public Entrepreneurship and Innovation Based on Machine Learning

Shi Mengyi[1] Shen Yuncong[2] Li Jiaojiao[1] Zhang Siwei[1] Xu Mengyu[1]
[1]Document and Information Center, Fudan University, Shanghai 200433
[2]Department of Mathematics, Fudan University, Shanghai 200433

Abstract: [Purpose/significance] Based on the open data related to scientific and technological innovation, this paper establishes the evaluation system and index for public entrepreneurship and innovation of the cities in China. **[Method/process]** By the machine learning algorithm, this paper established the Double-Innovation rank of domestic cities, and analyzed the Double-Innovation ability of the cities through various indicators and urban agglomeration characteristics. **[Result/conclusion]** The results show that the level of Double-Innovation in our country is closely related to economic resources, talent pool and policy environment. Beijing continues to lead Chinese dual-development, and the regional central cities are closely following it, these cities have their own advantages, and the development gap among them is small. The shortcomings of each city are also very clear. How to make reasonable countermeasures against their respective shortcomings, make up for the weaknesses, is the basis for the continuous development of urban innovation.

Keywords: double-innovation public entrepreneurship and innovation evaluation system double-innovation index